

Sensitivity of Multiview 3D Point Cloud Reconstruction to Compression Quality and Image Feature Detectability

*Ke Gao¹, *Shizeng Yao¹, Hadi AliAkbarpour¹, Sanjeev Agarwal², Guna Seetharaman³, Kannappan Palaniappan¹

¹University of Missouri-Columbia, MO, USA

²US Army, NVESD, Ft Belvoir, VA, USA

³US Naval Research Laboratory, Washington D.C., USA

kegao, shizeng.yao@mail.missouri.edu, aliakbarpourh, palaniappank@missouri.edu

sanjeev.agarwal.civ@mail.mil

guna.seetharaman@nrl.navy.mil

Abstract—In this paper we evaluate the quality of feature detection and 3D reconstruction on a Wide Area Motion Imagery (WAMI) sequence with increasing JPEG compression ratio. Feature detection is critical for computer vision tasks such as 3D reconstruction. For some 3D reconstruction approaches, the quality of a 3D model relies upon consistent detection of the same feature points over consecutive frames in an image sequence. Since the performance of feature detectors is highly sensitive to compression artifacts, we evaluate the influence of image quality on feature detection accuracy. Many datasets (e.g. WAMI) use JPEG compression to decrease the data storage and network bandwidth utilization while attempting to preserve image quality by adaptively adjusting the compression ratio. Consequently, it is important to understand the impact of JPEG compression on the quality of feature detection in 2D space and the subsequent 3D reconstruction results. We design and perform two evaluation procedures on the WAMI sequence. We use structure tensor to detect feature points on an image sequence with increasing JPEG compression ratio (10:1, 15:1, 20:1, 30:1, 40:1, 100:1, and 150:1). Compression ratio of 10:1 is used as the baseline (groundtruth). First we compare the feature points from images of different qualities with the groundtruth features and evaluate them on pixel level in 2D space. After that, a 3D model in the form of point cloud is generated from each set of feature points and compared with the groundtruth point cloud. We provide quantitative and visualized results for the evaluation.

Index Terms—JPEG compression, feature detection, 3D reconstruction, evaluation

I. INTRODUCTION

Image quality is of critical importance to computer vision tasks such as image registration, image retrieval, object recognition, and autonomous driving. Many computer vision applications are performed on smart handheld devices and unmanned aerial vehicles, which may not be equipped with the adequate computational resources to operate these algorithms in real time. The deployment of fast wireless transmission networks enables computer vision software to run on small devices by transmitting a large amount of data to remote servers.

Additionally, many computer vision algorithms are applied on large size images such as aerial imagery. Aerial imagery offers a new emerging challenge for computer vision applications for self-driving vehicles and drone systems. Processing aerial imagery involves enormous computation. Therefore, reducing power resources and bandwidth utilization become a key issue.

To address this problem, it is common to compress images in formats such as JPEG. JPEG has been a commonly used image compression standard for digital images since it was introduced decades ago [1]. JPEG compression is based on the Discrete Cosine Transform (DCT) [2] which is a lossy image compression method. The adjustable JPEG compression ratio allows a tradeoff between the image quality and image size. By adaptively tuning the compression ratio, many datasets attempt to preserve the image quality while decreasing the data storage. Digital images contain statistical properties and some details will inevitably be sacrificed by lossy compression. In certain scenarios, the compressed image quality is good enough if the difference between the compressed image and the original image is tolerable [3] [4]. However, compressing images may have a large impact on the performance of some computer vision tasks such as 2D feature detection. 3D reconstruction may also be influenced by image quality because feature detection is critical for 3D reconstruction [5]. Consequently, it is important to analyze the sensitivity of feature detection and 3D reconstruction to image compression quality. In this work, we perform two evaluation procedures and conduct experiments on aerial imagery with increasing JPEG compression ratios (10:1, 15:1, 20:1, 30:1, 40:1, 100:1, and 150:1). We provide quantitative results for the evaluation.

II. WIDE AREA MOTION IMAGERY

To analyze the sensitivity of feature detection and 3D reconstruction to image compression quality, we conduct experiments on one of the Wide Area Motion Imagery (WAMI) [6] [7] sequences (WAMI-ABQ215) that was collected by a fixed-wing aircraft with on-board INS sensors flying over the down-

*Ke Gao and Shizeng Yao contributed equally to this work.



Fig. 1. Sample frames of original size (6600×4400) from the WAMI-ABQ215 dataset. The frames are evenly separated in time.

town area in a city. Sample images are shown in Figure 1. The flight trajectory was a full orbit with the radius of $2500m$. The high resolution images of 6600×4400 pixels were captured airborne while the camera was locked (by a gimbal) to always look at the center of the scene. The typical distance between the centers of the camera and the scene was $\sim 3000m$. The WAMI-ABQ215 sequence consists of 215 RGB images. The total size of the uncompressed video is approximately 19 GB in total. We generated 7 set of image sequences from the original data using increasing JPEG compression ratio (10:1, 15:1, 20:1, 30:1, 40:1, 100:1, and 150:1). The compression ratio of 10:1 is considered as the baseline (ground truth) as JPEG typically achieves ten times compression with little perceptible loss in image quality [8]. The zoomed image of size 100×100 for each compression ratio are shown in Figure 2. There is little perceptible differences from compression ratio 10:1 to 40:1 despite the fact that the image size decreases significantly as the compression ratio increases. The compression artifacts are clearly visible on compression ratio 100:1 and 150:1. The data size and JPEG quality of 7 image sequences with increasing compression ratio are shown in Table I.

JPEG Compression Ratio	JPEG Quality	Dataset Size (RGB)	Image Size (RGB)
Uncompressed	100	~ 19 GB	84.60 MB
10:1 (default)	95 (default)	1999.50 MB	9.30 MB
15:1	92	1300.75 MB	6.05 MB
20:1	83	937.40 MB	4.36 MB
30:1	64	625.65 MB	2.91 MB
40:1	45	447.20 MB	2.08 MB
100:1	10	169.85 MB	0.79 MB
150:1	4	105.35 MB	0.49 MB

TABLE I
SIZE OF IMAGES WITH DIFFERENT JPEG COMPRESSION RATIO.

To generate the compressed images, we used the JPEG compression functions from both Matlab and OpenCV C++. The compression qualities from both methods are identical. The elapsed time for image compression are shown in Table II. The built-in JPEG compression function in OpenCV C++ is

approximately five times the speed of its Matlab counterpart.

JPEG Compression Ratio	Compression Time (Matlab)	Compression Time (OpenCV C++)
10:1 (default)	261.65 s	52.18 s
15:1	250.63 s	48.81 s
20:1	239.37 s	46.31 s
30:1	224.71 s	41.58 s
40:1	215.93 s	38.83 s
100:1	195.61 s	32.96 s
150:1	191.54 s	31.23 s

TABLE II
ELAPSED TIME FOR IMAGE COMPRESSION.

III. STRUCTURE TENSOR FOR 2D FEATURE DETECTION

We use the multi-scale 2D structure tensor [9] to detect 2D feature points in the compressed images with different JPEG compression ratios. First, a smoothed image is generated by convolving the input grayscale image with a Gaussian filter. After that, the first derivatives (horizontal and vertical) of the smoothed image $\mathbf{I}$ are computed as follows.

$$\nabla \mathbf{I}(x, y) = \begin{bmatrix} \frac{\partial \mathbf{I}}{\partial x} \\ \frac{\partial \mathbf{I}}{\partial y} \end{bmatrix} \quad (1)$$

Then three matrices corresponding to the outer products of the first derivatives are computed as the 2D structure tensor.

$$\mathbf{J}_{2D} = \nabla \mathbf{I} (\nabla \mathbf{I})^T = \begin{bmatrix} \left(\frac{\partial \mathbf{I}}{\partial x}\right)^2 & \frac{\partial \mathbf{I}}{\partial x} \frac{\partial \mathbf{I}}{\partial y} \\ \frac{\partial \mathbf{I}}{\partial x} \frac{\partial \mathbf{I}}{\partial y} & \left(\frac{\partial \mathbf{I}}{\partial y}\right)^2 \end{bmatrix} \quad (2)$$

The eigenvalues of $\mathbf{J}_{2D}$ are defined as:

$$\lambda_{1,2} = \frac{1}{2} (\mathbf{J}_{2D}(1,1) + \mathbf{J}_{2D}(2,2)) \quad (3)$$

The Beltrami scalar interest measure matrix can be computed as follows.

$$\begin{aligned} Beltrami(\mathbf{I}) &= 1 + (\lambda_1 + \lambda_2) + \lambda_1 \lambda_2 \\ &= 1 + \text{trace}(\mathbf{J}_{2D}) + \det(\mathbf{J}_{2D}) \end{aligned} \quad (4)$$

To extract multi-scale 2D feature points, a set of Gaussian filters of different sizes are used to convolve with the input image, which result in Beltrami matrices of multiple scales. And then a response matrix is calculated by measuring the maximum for each pixel across scales. Finally, non-maximum suppression is applied to the response matrix to extract the features. As can be seen in Figure 3, the detected feature points (red markers) are concentrated on the corners and edges of the buildings in the scene.

IV. MULTIVIEW 3D POINT CLOUD RECONSTRUCTION

To evaluate the quality of 3D point cloud reconstruction on the WAMI sequence with various JPEG compression ratio, we developed a video-based 3D reconstruction tool (VB3D) that uses multiple camera views and a novel volumetric voting-based surface likelihood map estimation scheme. VB3D can reconstruct city-scale 3D models of a few square kilometers. VB3D uses a salient feature detection based vote accumulation technique without reliance upon feature matching between

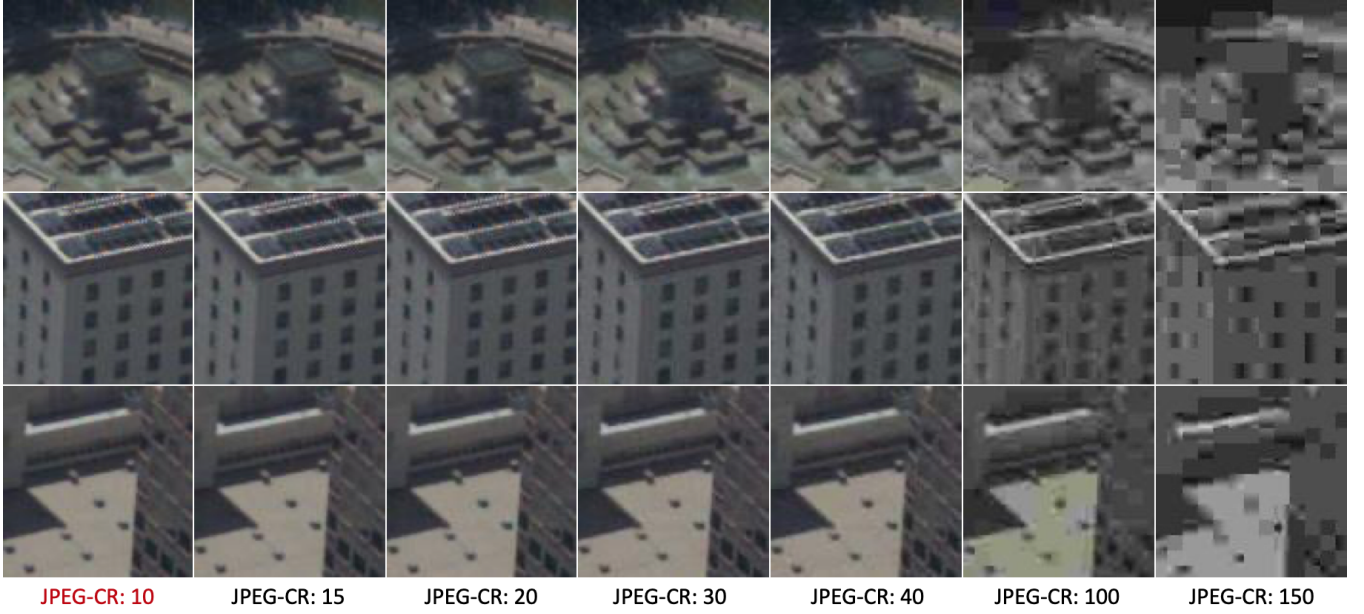


Fig. 2. Zoomed images (100×100) from the first frame in the WAMI-ABQ215 sequence with increasing JPEG compression ratio (10:1, 15:1, 20:1, 30:1, 40:1, 100:1, and 150:1). JPEG compression ratio 10:1 is used as the baseline for the subsequent experiments.

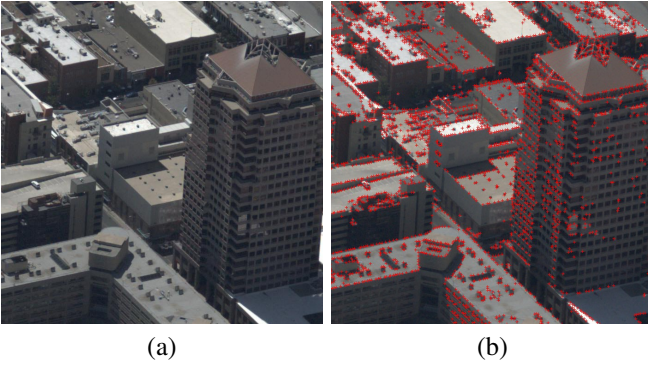


Fig. 3. Example of 2D feature point detection. (a) input image patch. (b) feature points detected by 2D structure tensor.

views to construct a point cloud representation of the 3D geometry of the objects in the scene. The workflow of VB3D is illustrated in Figure 4.

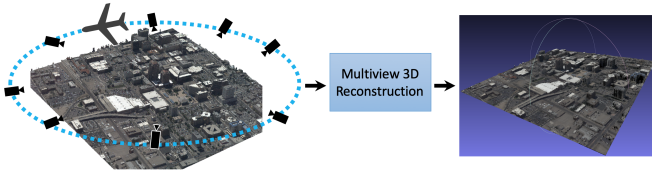


Fig. 4. Workflow of the multiview 3D point cloud reconstruction.

V. EXPERIMENTAL RESULTS

A. 2D Feature Detection

First, we plot the number of 2D feature points detected per frame over the image sequence with different JPEG compres-

sion ratios. As shown in Figure 5 (a), the number of feature points are similar from compression ratio 10:1 to 40:1. Starting from compression ratio 100:1, the number of features increases dramatically because the compression artifacts produces false positive features that do not have distinctive characteristics from its immediate neighborhood. We also compute the Dice coefficient of feature points per frame by comparing to the ground truth. Note that the feature points detected on images with compression ratio 10:1 are regarded as the ground truth. Having a detection feature mask and a ground truth feature mask where each feature point occupies only one pixel, the Dice coefficient can be calculated using the equation below.

$$Dice = \frac{2|Mask_{detect} \cap Mask_{gt}|}{|Mask_{detect}| + |Mask_{gt}|} \quad (5)$$

$$= \frac{2TP}{2TP + FP + FN}$$

where TP is the true positive points, FP is false positive points, and FN is false negative points. The higher the Dice coefficient, the more similar the detected features are compared to the ground truth. As can be seen in Figure 5 (b), the Dice coefficient of images over the whole sequence declines as the JPEG compression ratio increases. The TP , FP , and FN points are visualized in Figure 6 where each category is marked in different colors. The false positive features resulted from the compression artifacts can be clearly observed in Figure 6(f).

B. 3D Reconstruction

We reconstruct 3D point clouds by applying the multiview 3D reconstruction method introduced in Section IV to image sequences with increasing JPEG compression ratios and 2D feature points detected on them. The reconstructed 3D point clouds are shown in Figure 8. The 3D point cloud from JPEG

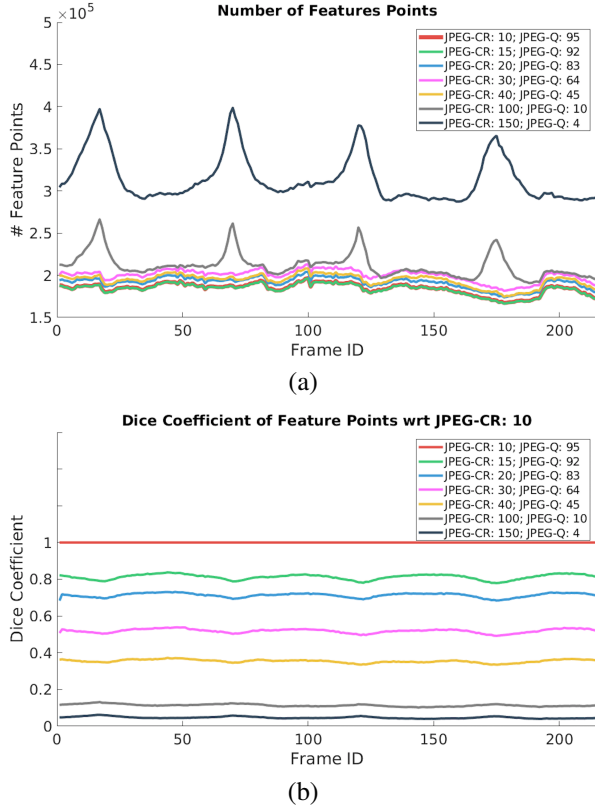


Fig. 5. Evaluation of 2D feature detection on WAMI-ABQ215 dataset (215 frames) with increasing JPEG compression ratios. (a) number of feature points detected per frame over the image sequence. (b) Dice coefficient of feature points per frame compared with the baseline (JPEG compression ratio: 10).

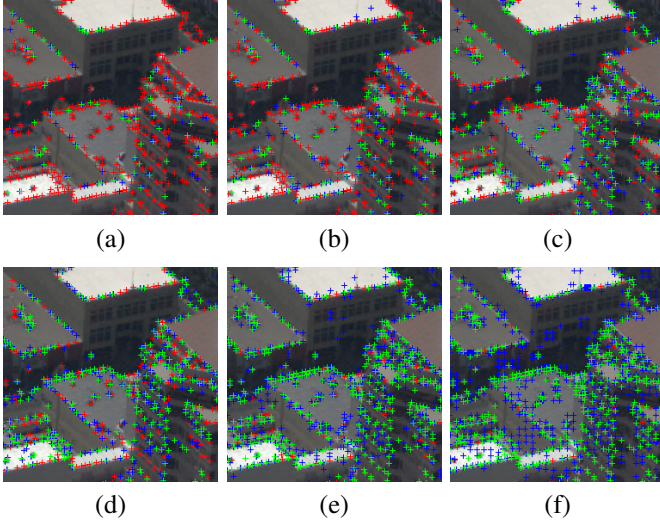


Fig. 6. Feature detection comparison. Red color corresponds to the true positive points, green color corresponds to the false negative points, and blue color corresponds to the false positive points. (a) compression ratio 10:1 vs 15:1. (b) compression ratio 10:1 vs 20:1. (c) compression ratio 10:1 vs 30:1. (d) compression ratio 10:1 vs 40:1. (e) compression ratio 10:1 vs 100:1. (f) compression ratio 10:1 vs 150:1.

compression ratio 10:1 is used as the baseline (ground truth). 90% accuracy and 1m completeness [10] [11] of each reconstructed point cloud are measured for quantitative evaluation. 90% accuracy is distance d such that 90% of the reconstruction is within d from the ground truth model. 1m completeness is the percentage of the ground truth model that is within 1m from the reconstruction. The quantitative evaluation results are presented in Table III. For JPEG compression ratio 15:1, the number of point clouds are close to the baseline and 90% of the reconstructed 3D points are within 0.6m from the baseline model. 98.68% of the 3D points in the baseline model are within 1m from the reconstruction. JPEG compression ratio 20:1, 30:1, and 40:1 generate reconstructions of similar quality in terms of number of 3D points and 90% accuracy. 1m completeness is approximately the same (97%) as well. Starting from compression ratio 100:1, the reconstruction models are highly different than the ground truth, which lead to high 90% accuracy and low 1m completeness.

Besides, we calculate the differences between each reconstructed point cloud model and the ground truth. The comparisons are visualized in Figure 7 where magenta color shows the 3D points from the ground truth model, green color shows the 3D points from the reconstructed model, and 3D points common to both are shown in white color. For compression ratios lower than 40:1, there is still a reasonable amount of shared points between the reconstruction model and the baseline. As the compression ratio increases, more noises are created especially on top of the buildings in the scene.

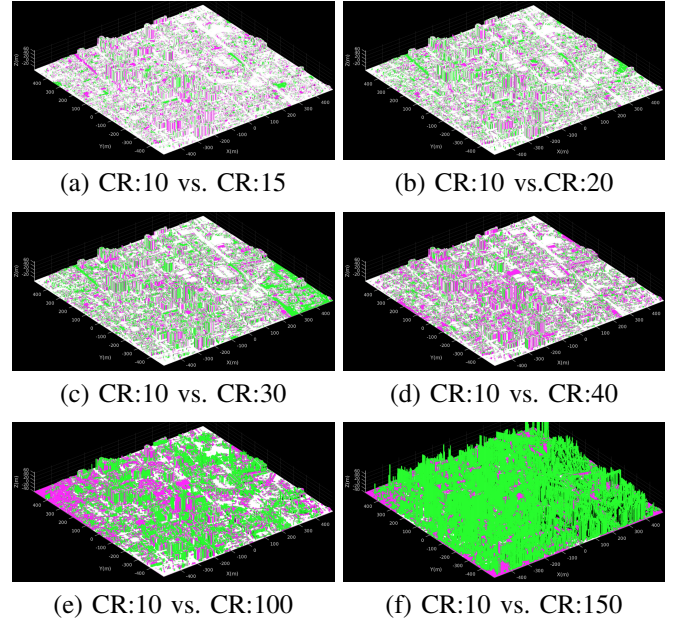


Fig. 7. Evaluation of 3D point cloud reconstruction on WAMI-ABQ215 dataset with increasing JPEG compression ratios compared with the baseline (JPEG-CR: 10). The voxels from the baseline point cloud are shown in magenta. The voxels from the reconstructed point cloud (JPEG-CR: 15, 20, 30, 40, 100, 150) are shown in green. The shared voxels are shown in white color.

JPEG Compression Ratio	JPEG Quality	Dataset Size (RGB)	Total Number of 2D Feature Points	Number of 3D Points Reconstructed	Point Cloud 90% Accuracy	Point Cloud 1m Completeness
10:1 (default)	95 (default)	1999.50 MB	39,609,038	~18.1 Million	0 m	100%
15:1	92	1300.75 MB	39,490,448	~18.1 Million	0.06 m	98.68%
20:1	83	937.40 MB	41,135,613	~18.1 Million	0.50 m	97.36%
30:1	64	625.65 MB	43,002,273	~18.1 Million	0.50 m	97.79%
40:1	45	447.20 MB	41,767,936	~18.2 Million	0.50 m	97.75%
100:1	10	169.85 MB	45,487,752	~17.4 Million	3.06 m	90.62%
150:1	4	105.35 MB	67,635,776	~34.8 Million	36.10 m	76.74%

TABLE III

QUANTITATIVE COMPARISONS OF 3D RECONSTRUCTION RESULTS WITH DIFFERENT JPEG COMPRESSION RATIOS.

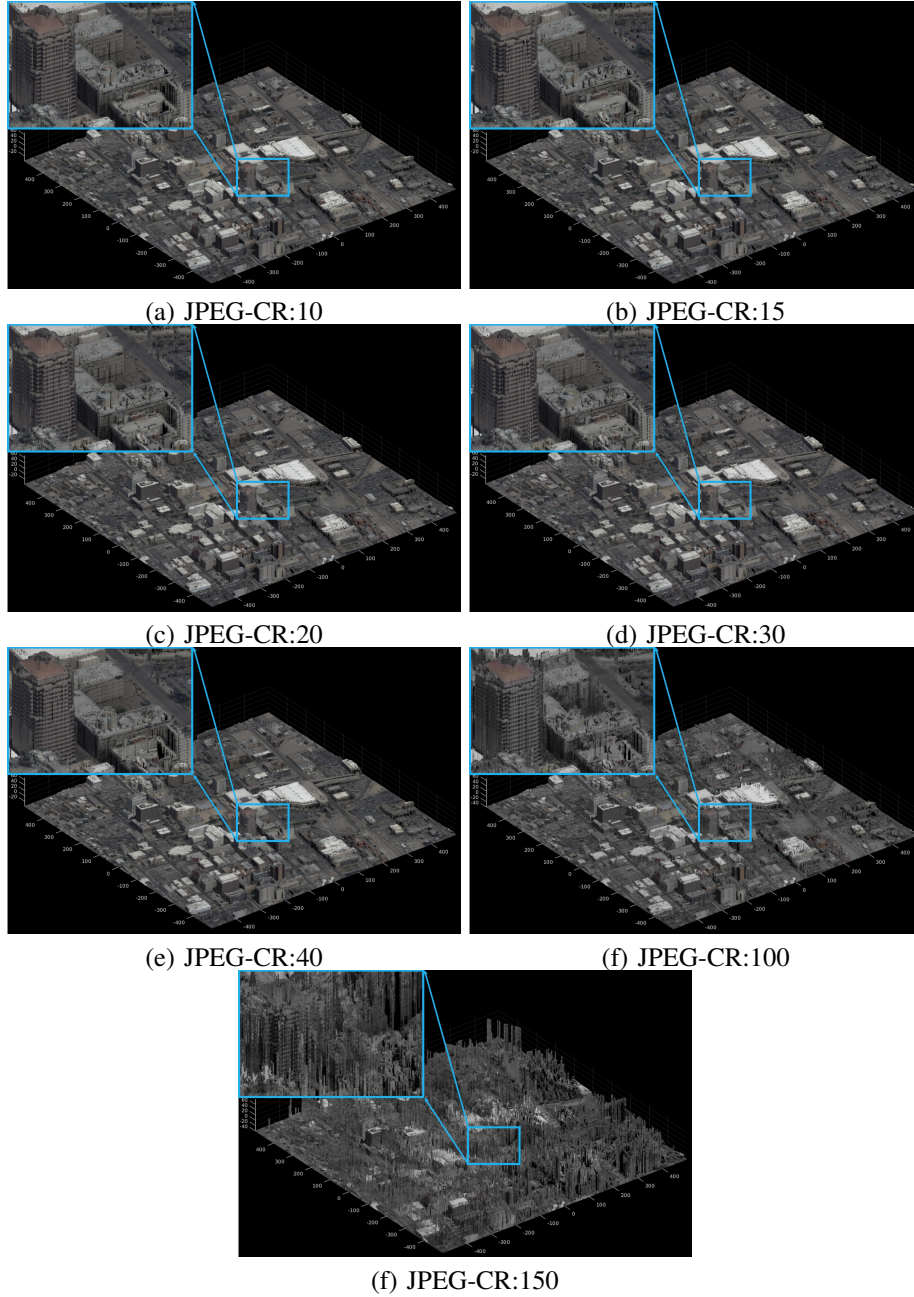


Fig. 8. 3D point cloud reconstruction results generated from image sequences with increasing JPEG compression ratios.

VI. CONCLUSIONS

In this work, we analyze the sensitivity of 2D structure tensor feature detection and multiview 3D point cloud reconstruction to image compression quality. We generate image sequences of increasing JPEG compression ratios (10:1, 15:1, 20:1, 30:1, 40:1, 100:1, and 150:1) on a Wide Area Motion Imagery (WAMI) dataset. Little perceptible difference can be observed for JPEG compression ratio 10:1, 15:1, 20:1, 30:1, and 40:1. High perceptible loss in image quality exists for compression ratio 100:1 and 150:1. We extract 2D feature points using structure tensor on the compressed images and present the evaluation results. The number of detected 2D feature points increases as the JPEG compression ratio increases, but the number of true positive points decreases because of compression artifacts. We also produce 3D reconstruction results using the compressed images and the 2D feature points detected on them. For JPEG compression ratio 10:1, 15:1, 20:1, 30:1, and 40:1, the reconstructed point clouds contain similar amount of 3D points. The reconstruction qualities are also similar in terms of 90% accuracy and 1m completeness. However, the 3D point clouds generated from images with compression ratio 100:1 and 150:1 are much less satisfactory both visually and quantitatively.

ACKNOWLEDGMENT

This work was partially supported by awards from U.S. Army Research Laboratory W911NF-1820285 and Army Research Office DURIP W911NF-1910181. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Government or agency thereof.

REFERENCES

- [1] G. Hudson, A. Léger, B. Niss, I. Sebestyén, and J. Vaaben, "JPEG-1 standard 25 years: past, present, and future reasons for a success," *Journal of Electronic Imaging*, vol. 27, no. 4, p. 040901, 2018.
- [2] N. Ahmed, "How I came up with the discrete cosine transform," *Digital Signal Processing*, vol. 1, no. 1, pp. 4–5, 1991.
- [3] K. R. Rao and P. Yip, *Discrete cosine transform: algorithms, advantages, applications*. Academic press, 2014.
- [4] M. Siddeq and M. Rodrigues, "A novel image compression algorithm for high resolution 3D reconstruction," *3D Research*, vol. 5, no. 2, p. 7, 2014.
- [5] K. Gao, H. AliAkbarpour, K. Palaniappan, and G. Seetharaman, "Evaluation of feature matching in aerial imagery for structure-from motion and bundle adjustment," in *Geospatial Informatics, Motion Imagery, and Network Analytics VIII*, vol. 10645. International Society for Optics and Photonics, 2018, p. 106450J.
- [6] R. Porter, A. M. Fraser, and D. Hush, "Wide-area motion imagery," *IEEE Signal Processing Magazine*, vol. 27, no. 5, pp. 56–65, 2010.
- [7] H. AliAkbarpour, K. Palaniappan, and G. Seetharaman, "Parallax-tolerant aerial image georegistration and efficient camera pose refinement without piecewise homographies," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4618–4637, 2017.
- [8] R. F. Haines and S. L. Chuang, "The effects of video compression on acceptability of images for monitoring life sciences experiments," 1992.
- [9] F. Bunyak, K. Palaniappan, S. K. Nath, and G. Seetharaman, "Flux tensor constrained geodesic active contours with sensor fusion for persistent object tracking," *Journal of Multimedia*, vol. 2, no. 4, p. 20, 2007.
- [10] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, 2006, pp. 519–528.
- [11] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1362–1376, 2009.